* here a list of all basic machine learning terms in 22 minutes artificial intelligence refers to the capability of machines to perform tasks that typically require human intelligence this can include understanding language recognizing images solving problems or making decisions AI aims to mimic human cognitive functions through various techniques including machine learning but not all AI is machine learning for example rule-based systems can use predefined logical rules to analyze medical data and provide diagnostic
* recommendations without needing to learn from data patterns typical chess playing engines would be considered AI but not machine learning because they follow specific rules in search algorithms and don't always learn from data machine learning is a branch of artificial intelligence that enables computers to learn from data and improve their performance on tasks over time without being explicitly programmed for each task in machine learning algorithms identify patterns and relationships within data making predictions or
* decisions based on new unseen information for example a spam filter in an email system uses machine learning to identify and block spam emails it is trained on thousands of examples of both spam and non-spam emails learning which words phrases or patterns are typically found in spam messages over time it can accurately flag new emails of spam or legitimate based on these learned patterns even if the specific content of each new email varies in many ways this is similar to how animals and humans learn to recognize patterns over time
* after seeing many examples of something for example a human child might not be able to tell the difference between a cat and a dog but after years of having someone point out cats and dogs it will learn to recognize the features that determine what a cat cat and a dog is an algorithm is a set of well-defined instructions or rules that a computer follows to solve a problem or perform a task algorithms are used in almost every aspect of computing from sorting lists and searching data to more complex processes like encryption and data
* analysis they provide step-by-step procedures to achieve a specific goal efficiently think for example of a step-by-step recipe like this sandwich making algorithm an example is dy stress algorithm used in mapping applications to find the shortest path between two points by systematically evaluating possible paths backstress algorithm helps determine the quickest route for navigation which is at the base of most navigation apps like Google Maps data is information that can be collected analyzed and used to make decisions
* predictions or provide insights than spreadsheets in Computing and machine learning data typically consists of numbers text images or any form of input that can be processed by algorithms for example customer purchase histories are a type of data that e-commerce companies analyze to recommend products likely to interest each user another example is weather data which includes temperature humidity and wind speed measurements this data is used to predict future weather patterns in the case of images data refers to a list of pixel
* intensities and possibly colors used by image recognition algorithms in the case of text Data could simply be a list of words and a text in their frequencies data can come in many forms a model in machine learning is a mathematical representation that is trained to recognize patterns in data and make predictions or classifications based on those patterns the most common type of model is simply a mapping function between a an input and an output in linear regression for example the model is simply the equation of the final
* regression line in its simplest form we might have a model that predicts a linear relationship between square footage of the house and the price of the house for example if we plot all house prices and their square footage against each other we might find that on average each additional square foot adds $200 to the house price the number 200 comes from the fitting of a line to the data which is now our train model the train model is the intersection and slope of the line the slope being 200 model fitting also called training or
* learning is the process of adjusting a model's parameters to find the best match between the model's predictions and the actual data if you think of linear regression model fitting would be trying out different lines until you find the line with the best fit training data is a carefully selected subset of data used to teach machine learning models how to make predictions it consists of input examples paired with their correct outputs allowing the model to learn patterns and relationships for instance in an email spam filter the
* training data would include thousands of emails labeled as either spam or not spam teaching the system to recognize the characteristics of unwanted messages similarly for an image recognition system that identifies cats and dogs the training data would contain numerous images labeled as either cat or dog helping the model learn the visual patterns that Define what cats and dogs look like test data or test set is a separate collection of data used to evaluate how well a machine learning model performs on examples it hasn't
* seen during training like training data it includes both inputs and their correct answers but these examples are kept completely separate from the training process this testing process helps verify whether the model has truly learned to make good predictions rather than just memorizing its training examples importantly the test and training data are separated randomly before beginning the modeling process so that the model can never see the test data in any way before running the final test any inadvertent inclusion
* of even parts of the test data in model training is called Data leakage supervised learning is a foundational approach in machine learning where models learn from labeled examples meaning the true outcomes or targets are known and provided much like a student learning from problems with their answers provided each example in the training data includes both of the input and the correct output allowing the model to learn the relationship between them for instance an image recognition system would train on images that have
* been pre-labeled with their contents such as dog or cat this is arguably the most common type of machine learning probably making up around 70% of machine learning applications unsupervised learning is a type of machine learning where models learn to find patterns and structure in data without being given labeled examples or correct answers rather than being taught what to look for these algorithms discover natural groupings and relationships within the data on their own for example an unsupervised learning algorithm might
* analyze customer purchase data to identify groups of customers with similar buying habits or examine social media posts to discover trending topics all without being told in advance what patterns to look for this approach is particularly valuable when we want to explore data to uncover hidden patterns but don't know exactly what we're looking for No Labels or outcomes are provided to the model during training reinforcement learning is a newer branch of machine learning that has recently been accepted as a third main branch of
* machine learning and has gain momentum in the late 2010s particularly with the success of deep Minds chess engine alphago in 2016 it's distinct from both supervised and unsupervised learning because it operates on a fundamentally different principle instead of learning from pre-labeled examples supervised or finding patterns in unlabeled data unsupervised it learns from interaction and feedback unlike supervised learning where examples have clear right answers reinforcement learning is more like training a pet the agent learns through
* trial and error getting rewarded for good decisions and penalized for poor ones for example a reinforcement learning algorithm can learn to play chess by playing thousands of games against itself receiving positive rewards for winning moves and negative rewards for losing ones this approach is particularly powerful for tasks involving sequential decision-making like gam playing robotic control or optimizing business strategies when there are no clear labels but an idea of what is a good or bad outcome many basic
* machine learning courses don't cover reinforcement learning as a basic machine learning Branch but as an advanced topic since it is still fairly Niche a feature also called a predictive variable input variable independent variable or attribute is a specific piece of information or characteristic used as input for a machine learning model essentially it's any measurable property that helps the model make predictions for example in a house price prediction model features might include the square footage number of bedrooms
* location and age of the house for an email spam detector features could include the number of capitalized words the number of URLs in the text or whether the sender is in your contacts the selection and Engineering of relevant features sometimes called feature extraction or feature design is often crucial to a model success as they need to capture the important aspects of the data that relate to the prediction task feature engineering is the process of creating new more informative features from existing raw data to
* improve a model's performance feature engineering involves using domain knowledge and creativity to transform or combine original features into more meaningful ones for for example instead of just using raw date values you might create features like day of the week or is holiday which will probably explain fluctuations of sales much better good feature engineering often makes the difference between an average model and an excellent one as it helps the model focus on the most relevant patterns in the data feature scaling also called
* normalization or standardization is the process of transforming numeric features to a similar scale typically to prevent features with larger ranges from dominating the learning process for example here the numbers for salary are much larger than those for age and dominate the model fitting common scaling methods include minmax normalization thus scaling to a 0 to one range as seen here or standardization transforming to mean zero and standard deviation one proper scaling is particularly important for many algorithms like gradient descent and
* neural networks which can perform poorly or converge slowly when features are on vastly different scales dimensionality refers to the number of features also called Dimensions variables or attributes in a data set for example in a house price prediction model if each house is described by square footage number of bedrooms location age number of bathrooms and distance from the city center the data has six dimensions High dimensional data having many features can pose unique challenges often called The Curse of dimensionality as
* Dimensions increase data becomes more sparse and patterns become harder to find much like trying to find a needle in an increasingly large Hast stack this is why dimensionality reduction techniques are often crucial in machine learning helping to compress many features into a smaller set while preserving important information feature engineering feature scale and dimensionality reduction are all part of data pre-processing along with other techniques a Target also called the dependent variable output variable
* response variable or label is what a machine learning model is trying to predict based on the features for example in a house price prediction model the target would be the actual sale price of the house while in an email spam detector the target would be whether an email is Spam or not spam in supervised learning the training data must include both features and their corresponding Target values allowing the model to learn the relationship between them an instance also called a sample example record data point or observation
* is a single complete unit of data that includes all features and in supervised learning it's Target value in this example it's one person with their name age income and marital status for house prediction it might be a single house with all its characteristics like square footage and price a typical machine learning data set consists of many such instances which together form the training or test data think of an instance as one row in a data table or spreadsheet with the columns being the feature features and the target the
* entire table would be called your data set a label also called a class Target value ground truth or correct answer is the known correct output associated with an instance in supervised learning it is the value that the target variable takes for each instance in an image recognition system where the target variable is the type of animal in the picture the label is the actual animal's name like cat or dog for each image labels are crucial for training supervised learning models as they provide the right answers that the model
* learns from to obtaining accurate labels often require significant human effort such as experts manually categorizing thousands of examples this process is called labeling which often is a major bottleneck in supervised learning well-labeled data is a hot commodity and many creative ways exist to generate it including crowdsourcing model complexity refers to how sophisticated a machine learning model is in terms of its ability to capture patterns in the data a more complex model has more parameters and can learn more complicated
* relationships like a neural network with many layers conversely a simple model has fewer parameters and can only capture basic patterns like a linear regression finding the right level of complexity is crucial too simple and the model fails to capture important patterns which is called underfitting too complex and it learns to fit to noise in the training data rather than true patterns also called overfitting a simple way to think about model complexity is by thinking about the polinomial order of a regression line a
* simple linear regression only has to estimate the intercept and the slope of the line so two parameters a quadratic regression has to estimate the intercept and two parameters and so on each polinomial can potentially fit more complicated data this relationship between polinomial order and complexity provides a clear example of the trade-off between a model's ability to capture complex patterns and its risk of fitting to noise bias in terms of model complexity refers to how limited or inflexible a model's assumptions are
* about the underlying patterns in the data a model with high bias like a linear regression makes strong simple assumptions in this case that the relationship is purely linear as we increase the polinomial order the bias decreases a second order polinomial has more flexibility to fit curves low bias means fewer built-in assumptions about the data structure this doesn't mean lower bias is always better a very high order polinomial might have such low bias that it fits the training data perfectly but fails to generalize well
* leading to overfitting variance refers to how much a model's predictions would change if it were trained on different subsets of the training data a model with high variance is very sensitive to small changes in the training data producing significantly different predictions when trained on slightly different data sets models with low variance like linear regression produce more consistent predictions across different training sets High variance often indicates overfitting where the model is learning the random noise and the training data
* rather than the true underlying patterns there's typically a trade-off between bias and variance where reducing one tends to increase the other the bias variance trade-off is a fundamental Concept in machine learning that describes the tension between a model's ability to minimize bias and variance at the same time as model complexity increases bias typically decreases because the model can capture more complex patterns but variance increases because the model becomes more sensitive to changes in the training data
* conversely as model complexity decreases bias increases because the model makes more rigid assumptions but variance decreases because the model becomes more stable finding The Sweet Spot in this tradeoff is crucial the goal is to create a model that's complex enough to capture true patterns in the data but not so complex that it fits to noise this balance typically produces the best generalization to new data this concept is one of the most Central and important concepts of machine learning truly understanding this concept on all levels
* will make you a great data scientist and machine learning engineer noise refers to random variations or errors in data that don't represent true underlying patterns like random fluctuations in sensor readings or errors in data collection in machine learning we want to find the true patterns while ignoring this noise noise is what's left over after perfect fitting of the data with a perfect model capturing all the signal in the data overfitting occurs when a machine learning model learns the noise and random fluctuations in the training
* data rather than learning the true underlying patterns like a student who memorizes test answers without understanding the concepts an overfitted model performs well on training data but fails to generalize to new examples this typically happens when a model is too complex for the task or when it trains for too long on too little data causing it to mistake random noise for Meaningful patterns the model has high variance under fitting occurs when a machine learning model is too simple to capture the important patterns in the
* data resulting in poor performance on both training and test data like using a straight line to model clearly curved data an underfitted model makes oversimplified assumptions about the underlying patterns this typically happens when a model has high bias for example using a linear model to capture relationships that are clearly nonlinear one way to estimate bias and variance during training and thus avoid underfitting and overfitting before applying your model to real world data is validation validation is the practice
* of evaluating a model's performance on data it hasn't been trained on by setting aside a portion of the training data called the validation set to simulate how well the model will perform on new unseen data cross validation extends this concept by repeatedly training and validating the model on different splits of the data for example in five-fold Cross validation the data is divided into five parts and the model is trained five times each time using a different part as the validation set and the remaining parts for training this
* practice provides a more robust estimate of the model's True Performance and helps detect potential issues like overfitting or underfitting while validation sets are used during the model development process to make decisions about hyperparameters and model selection the test set is kept completely separate and used only once at the very end to evaluate the final model's performance using the test set repeatedly would risk overfitting to it regularization refers to techniques used to prevent overfitting by adding
* constraints or penalties that discourage a model from becoming too complex or fitting too closely to the train training data it keeps the model parameters small you can think of it as squeezing the regression lens so it doesn't become too wild the strength of the regularization is a hyperparameter too much regularization leads to underfitting a batch is a subset of training data that is processed together in a single step of model training rather than processing the entire data set at once for example instead of using
* all 10,000 training images simultaneously a model might process batches of 32 images at a time updating its parameters after each batch the batch size is an important typer parameter that affects training larger batches provide more stable parameter updates but require more memory while smaller batches update more frequently and can help the model Escape local Optima an iteration is a single pass through one batch of data leading to an update of the parameters of the model an Epoch is a complete pass through the
* entire training data set during model training this means each batch and thus each training example has been seen and learned from Once models typically need multiple epochs to learn effectively with each pass refining its understanding however too many epochs can lead to overfitting where the model starts memorizing the training data rather than learning General patterns these things only come into play for very large data sets that need to be split into batches small data sets are not split a parameter also called a
* model parameter or weight is a value that the model learns during training from the data unlike hyperparameters which are set before training begins finding the parameters of a model is the goal of the training process for example in a linear regression model the slope M and intercept B are parameters that the model adjusts to fit the data in more complex models like neural networks parameters include all the weights and biases that are automatically adjusted during training to minimize prediction errors weights and biases correspond to
* the slope and intercept of linear regression while a typical linear regression might have just a few parameters modern deep learning models can have millions or even billions of parameters each being fine-tuned through the training process to capture patterns in the data a hyperparameter is a configuration setting used to control the learning process set before training begins unlike model parameters which are learned during training examples include the learning rate batch size number of epochs or the number of layers in a
* neural network these are like the knobs and dials that data scientists adjust to optimize how a model learns finding the right hyperparameter values often requires experimentation as their optimal settings can vary significantly between different problems and data sets a cost function also called a loss function objective function or error function is a measure of how wrong a model's predictions are compared to the True Values it quantifies the cost or penalty of incorrect predictions for example in a house price prediction
* model the cost might be the average difference between predicted and actual prices so in a linear regression model as seen here we often use the mean squared error function that is the squared vertical distances of the data points from the regression line here that is the sum of all the red square areas the further the line from the actual data points the larger the error which we also call loss or cost the goal of training is to minimize this cost function like trying to achieve the lowest possible error score the specific
* choice of cost function significantly influences how the model learns and what kinds of Errors it prioritizes avoiding and can be considered another hyperparameter gradient descent is a fundamental optimization algorithm used to train machine learning models by iteratively adjusting model parameters to minimize errors it is one of the main methods for minimizing the cost function like a hiker trying to find the lowest point in a valley by always stepping in the steepest downhill Direction gradient descent calculates the direction in
* which the model's error decreases most rapidly and updates the parameters accordingly for each step it computes the gradient essentially the slope of the error with respect to each parameter then adjust these parameters in the opposite direction of the gradient using the learning rate to determine step size this process continues until the model reaches a minimum error or stops improving significantly interestingly a ball rolling down a mountain will behave the same way at each point only going in the direction of the steepest Ascent
* this is Nature's gradient descent but as you can imagine the ball can also get stuck in a local minimum like a depression on the mountain side instead of finding its way all the way down to the valley however a real ball in particular a heavy one has momentum which allows it to shoot over local depressions and keep going down the valley this inspired a variant of gradient descent called momentum based gradient descent Which is less likely to get stuck in local Minima the learning rate is a crucial hyperparameter that
* determines how much a model adjusts its parameters in response to errors during training like a student adjusting their understanding based on on feedback a model with a high learning rate makes large adjustments to its parameters after seeing each batch of data potentially learning quickly but risking overshooting optimal values conversely a model with a low learning rate makes smaller more cautious adjustments this can be more stable but might take longer to converge or get stuck in suboptimal Solutions finding the right learning
* rate is often critical for successful training too high and the model might never converge too low and training might take unnecessarily long evaluation is the process of measuring how well a machine learning model performs on data it hasn't seen during training using various metrics appropriate to the task for classification model evaluation might involve measuring accuracy precision recall or F1 score for regression model it might use mean squared error or R squar values this process typically involves both
* validation to tune the model during development and testing using a completely separate test set to get an unbiased estimate of final performance evaluation helps determine whether a model has truly learned use patterns or has just memorized the training data those were all basic machine learning terms in 22 minutes although I surely missed a bunch if I did please complain in the comments if you found this video helpful share it with someone who you think might also like it and get started on one of the tutorials in the
* description or on this very Channel also consider liking the video and subscribing to be notified about similar content in the future thanks for watching